Informational Divergence Approximations to Product Distributions

Jie Hou and Gerhard Kramer
Institute for Communications Engineering
Technische Universität München, 80290 Munich, Germany
Email: {jie.hou, gerhard.kramer}@tum.de

***Abstract*—The minimum rate needed to accurately approximate a product distribution in terms of an unnormalized informational divergence is shown to be a mutual information. The result follows by evaluating nonasymptotic results of Hayashi. An alternative and direct proof is given that extends to cases where the source distribution is unknown but the source entropy is known.**

I. Introduction

What is the minimal rate needed to generate a good approximation of a target distribution with respect to some distance measure? For example, to learn a system response, we might give inputs to the system and compute the output statistics. However, in computer simulations the inputs are only some approximations of the true distributions that are generated with random number generators. We would like to use a small number of bits to generate good approximations of a target distribution.

Wyner considered such a problem and characterized the smallest rate needed to approximate a *product* distribution accurately when using the *normalized* informational divergence as the distance measure between two distributions. The smallest rate is a Shannon mutual information [1]. Han-Verdú [2] showed that the same rate is necessary and sufficient to generate distributions arbitrarily close to an *information stable* distribution in terms of *variational distance*. Note that normalized informational divergence and variational distance are not necessarily larger or smaller than the other. Hayashi [3] studied resolvability using the *unnormalized* informational divergence and derived results for non-asymptotic cases that can be extended to asymptotic cases.

We show that the minimal rate needed to make the *unnormalized* informational divergence between a target product distribution and the approximating distribution arbitrarily small is the same Shannon mutual information as in [1], [2]. This result also follows from [3, Lemma 2] although it was not stated in [3]. Thus, our main contributions might be considered to be an alternative proof to that in [3], and we extend the proof to cases where the encoder has a non-uniform input distribution. In any case, we emphasize that the result implies the results in [1] and [2] when restricting attention to product distributions (in particular Theorem 6.3 in [1] and Theorem 4 in [2]).

The paper is organized as follows. In Section II, we state the problem. In Section III we state and prove the main result. Section IV discusses extensions.

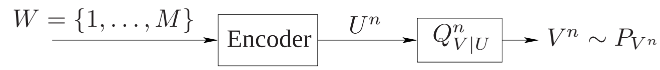

Fig. 1. Coding problem with the goal of making $P_{V^n} \approx Q_V^n$.

II. Preliminaries

Random variables are written with upper case letters and their realizations with the corresponding lower case letters. Superscripts denote finite-length sequences of variables/symbols, e.g., $X^n = X_1,\ldots,X_n$. Subscripts denote the position of a variable/symbol in a sequence. For instance, X_i denotes the i-th variable in X^n. A random variable X has probability distribution P_X and the support of P_X is denoted as supp(P_X). We write probabilities with subscripts $P_X(x)$ but we drop the subscripts if the arguments of the distribution are lower case versions of the random variables. For example, we write $P(x) = P_X(x)$. If the X_i, $i = 1,\ldots,n$, are independent and identically distributed (i.i.d.) according to P_X, then we have $P(x^n) = \prod_{i=1}^n P_X(x_i)$ and we write $P_{X^n} = P_X^n$. Calligraphic letters denote sets. The size of a set $\mathcal{S}$ is denoted as $|\mathcal{S}|$. We use $T_\epsilon^n(P_X)$ to denote the set of letter-typical sequences of length n with respect to the probability distribution P_X and the non-negative number ϵ [4, Ch. 3], [5], i.e., we have

$$T_\epsilon^n(P_X) = \left\{ x^n : \left| \frac{N(a|x^n)}{n} - P_X(a) \right| \le \epsilon P_X(a),\ \forall a \in \mathcal{X} \right\}$$

where $N(a|x^n)$ is the number of occurrences of a in x^n.

Consider the system depicted in Fig. 1. The random variable W with cardinality $M = 2^{nR}$ is *uniformly* distributed over $\{1,\ldots,M\}$ and is encoded to sequences U^n. V^n is generated from U^n through a memoryless channel $Q^n_{V|U}$ and has distribution P_{V^n}. A rate R is *achievable* if for any $\delta > 0$ there is a sufficiently large n and an encoder such that

$$D(P_{V^n}||Q_V^n) = \sum_{v^n \in \text{supp}(P_{V^n})} P(v^n) \log \frac{P(v^n)}{Q_V^n(v^n)} \qquad (1)$$

is less than δ. We wish to determine the smallest achievable rate.

III. Achievability

Theorem 1: For a given target distribution Q_V, the rate R is achievable if $R > I(V;U)$, where $I(V;U)$ is calculated with some joint distribution Q_{UV} that has marginal Q_V

and $|\text{supp}(Q_U)| \leq |\mathcal{V}|$. The rate R is not achievable if $R < I(V;U)$ for all Q_{UV} with $|\text{supp}(Q_U)| \leq |\mathcal{V}|$.

Proof: Suppose U and V have finite alphabets $\mathcal{U}$ and $\mathcal{V}$, respectively. Let Q_{UV} be a probability distribution with marginals Q_U and Q_V. Let $U^nV^n \sim Q_{UV}^n$, i.e., for any $u^n \in \mathcal{U}^n$, $v^n \in \mathcal{V}^n$ we have

$$Q(u^n, v^n) = \prod_{i=1}^{n} Q_{UV}(u_i, v_i) = Q_{UV}^n(u^n, v^n) \tag{2}$$

$$Q(u^n) = \prod_{i=1}^{n} Q_U(u_i) = Q_U^n(u^n) \tag{3}$$

$$Q(v^n) = \prod_{i=1}^{n} Q_V(v_i) = Q_V^n(v^n) \tag{4}$$

$$Q(v^n|u^n) = \prod_{i=1}^{n} Q_{V|U}(v_i|u_i) = Q_{V|U}^n(v^n|u^n). \tag{5}$$

Let $\mathcal{C} = \{U^n(w)\}_{w=1}^M$, where the $U^n(w), w = 1, \ldots, M$, are generated in an i.i.d. manner using Q_U^n. V^n is generated from $U^n(W)$ through the channel $Q_{V|U}^n$ (see Fig. 2). We have

$$P(v^n) = \sum_{w=1}^{M} \frac{1}{M} \cdot Q_{V|U}^n(v^n|u^n(w)). \tag{6}$$

Note that if for a v^n we have

$$Q_V^n(v^n) = \sum_{u^n \in \text{supp}(Q_U^n)} Q_U^n(u^n) Q_{V|U}^n(v^n|u^n) = 0 \tag{7}$$

then we have

$$Q_{V|U}^n(v^n|u^n) = 0, \text{ for all } u^n \in \text{supp}(Q_U^n). \tag{8}$$

This means $P(v^n) = 0$ and $\text{supp}(P_{V^n}) \subseteq \text{supp}(Q_V^n)$ so that $D(P_{V^n}||Q_V^n) < \infty$. We further have

$$\begin{aligned} \mathrm{E}\left[\frac{Q_{V|U}^n(v^n|U^n)}{Q_V^n(v^n)}\right] &= \sum_{u^n} Q_U^n(u^n) \cdot \frac{Q_{V|U}^n(v^n|u^n)}{Q_V^n(v^n)} \\ &= 1 \end{aligned} \tag{9}$$

The average informational divergence over all codebooks $\mathcal{C}$ is (recall that $P(w) = \frac{1}{M}, w = 1, \ldots, M$):

$$\begin{aligned} &\mathrm{E}[D(P_{V^n}||Q_V^n)] \overset{(a)}{=} \mathrm{E}\left[\log \frac{\sum_{j=1}^M \frac{1}{M} \cdot Q_{V^n|U^n}(V^n|U^n(j))}{Q_V^n(V^n)}\right] \\ &= \sum_w \frac{1}{M} \cdot \mathrm{E}\left[\log \frac{\sum_{j=1}^M Q_{V|U}^n(V^n|U^n(j))}{MQ_V^n(V^n)} \middle| W = w\right] \\ &\overset{(b)}{\leq} \sum_w \frac{1}{M} \cdot \mathrm{E}\left[\log\left(\frac{Q_{V|U}^n(V^n|U^n(w))}{MQ_V^n(V^n)} + \frac{M-1}{M}\right) \middle| W = w\right] \\ &\leq \sum_w \frac{1}{M} \cdot \mathrm{E}\left[\log\left(\frac{Q_{V|U}^n(V^n|U^n(w))}{MQ_V^n(V^n)} + 1\right) \middle| W = w\right] \\ &\overset{(c)}{=} \mathrm{E}\left[\log\left(\frac{Q_{V|U}^n(V^n|U^n)}{M \cdot Q_V^n(V^n)} + 1\right)\right] \end{aligned} \tag{10}$$

where

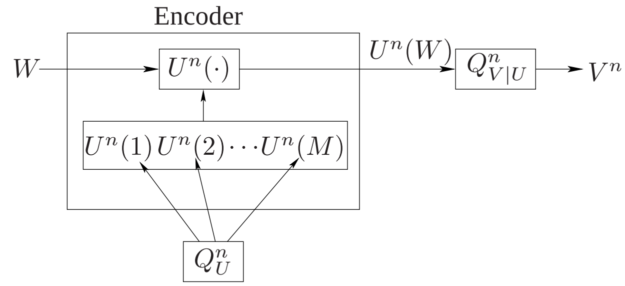

Fig. 2. The random coding experiment.

(a) follows by taking the expectation over $W, U^n(1), \ldots, U^n(M), V^n$;
(b) follows by the concavity of the logarithm and Jensen's inequality applied to the expectation over the $U^n(j), j \neq w$, and by using (9);
(c) follows by choosing $U^nV^n \sim Q_{UV}^n$.

Alternatively, we can make the steps (10) more explicit:

$$\begin{aligned} &\mathrm{E}[D(P_{V^n}||Q_V^n)] \\ &\overset{(a)}{=} \sum_{u^n(1)} \cdots \sum_{u^n(M)} \prod_{k=1}^{M} Q_U^n(u^n(k)) \\ &\sum_{v^n} \sum_{w=1}^{M} \frac{1}{M} \cdot Q_{V|U}^n(v^n|u^n(w)) \left[\log \frac{\sum_{j=1}^M Q_{V|U}^n(v^n|u^n(j))}{M \cdot Q_V^n(v^n)}\right] \\ &= \sum_w \frac{1}{M} \sum_{v^n} \sum_{u^n(w)} Q_{UV}^n(u^n(w), v^n) \\ &\sum_{k \neq w} \sum_{u^n(k)} \prod_{l \neq w} Q_U^n(u^n(l)) \left[\log \frac{\sum_{j=1}^M Q_{V|U}^n(v^n|u^n(j))}{M \cdot Q_V^n(v^n)}\right] \\ &\overset{(b)}{\leq} \sum_w \frac{1}{M} \sum_{v^n} \sum_{u^n(w)} Q_{UV}^n(u^n(w), v^n) \\ &\left[\log\left(\frac{Q_{V|U}^n(v^n|u^n(w))}{M \cdot Q_V^n(v^n)} + \sum_{j \neq w} \sum_{u^n(j)} \left[\frac{Q_{UV}^n(u^n(j), v^n)}{M \cdot Q_V^n(v^n)}\right]\right)\right] \\ &= \sum_w \frac{1}{M} \sum_{v^n} \sum_{u^n(w)} Q_{UV}^n(u^n(w), v^n) \\ &\left[\log\left(\frac{Q_{V|U}^n(v^n|u^n(w))}{M \cdot Q_V^n(v^n)} + \frac{M-1}{M}\right)\right] \\ &\leq \sum_w \frac{1}{M} \sum_{v^n} \sum_{u^n(w)} Q_{UV}^n(u^n(w), v^n) \\ &\quad \left[\log\left(\frac{Q_{V|U}^n(v^n|u^n(w))}{M \cdot Q_V^n(v^n)} + 1\right)\right] \\ &\overset{(c)}{=} \mathrm{E}\left[\log\left(\frac{Q_{V|U}^n(V^n|U^n)}{M \cdot Q_V^n(V^n)} + 1\right)\right]. \end{aligned} \tag{11}$$

We may write (10) or (11) as

$$\mathrm{E}\left[\log\left(\frac{Q_{V|U}^n(V^n|U^n)}{M \cdot Q_V^n(V^n)} + 1\right)\right] = d_1 + d_2 \tag{12}$$

where

$$d_1 = \sum_{(u^n,v^n)\in T_\epsilon^n(Q_{UV})} Q(u^n,v^n)\log\left(\frac{Q(v^n|u^n)}{M\cdot Q(v^n)}+1\right)$$

$$d_2 = \sum_{\substack{(u^n,v^n)\notin T_\epsilon^n(Q_{UV})\\(u^n,v^n)\in\text{supp}(Q_{UV}^n)}} Q(u^n,v^n)\log\left(\frac{Q(v^n|u^n)}{M\cdot Q(v^n)}+1\right).$$

Using standard inequalities (see [5]) we have

$$\begin{aligned} d_1 &\le \sum_{(u^n,v^n)\in T_\epsilon^n(Q_{UV})} Q(u^n,v^n)\log\left(\frac{2^{-n(H(V|U)-\epsilon)}}{M\cdot 2^{-n(H(V)+\epsilon)}}+1\right)\\ &\le \log\left(\frac{2^{-n(H(V|U)-\epsilon)}}{M\cdot 2^{-n(H(V)+\epsilon)}}+1\right)\\ &= \log\left(2^{-n(R-I(V;U)-2\epsilon)}+1\right)\\ &\le \log(e)\cdot 2^{-n(R-I(V;U)-2\epsilon)} \end{aligned} \tag{13}$$

and $d_1 \to 0$ if $R > I(V;U)+2\epsilon$ and $n\to\infty$. We further have

$$\begin{aligned} d_2 &\le \sum_{\substack{(u^n,v^n)\notin T_\epsilon^n(Q_{UV})\\(u^n,v^n)\in\text{supp}(Q_{UV}^n)}} Q(u^n,v^n)\log\left(\left(\frac{\nu_{V|U}}{\mu_V}\right)^n+1\right)\\ &\le 2|\mathcal{V}|\cdot|\mathcal{U}|\cdot e^{-2n\epsilon^2\mu_{UV}^2}\log\left(\left(\frac{\nu_{V|U}}{\mu_V}\right)^n+1\right) \end{aligned} \tag{14}$$

where

$$\mu_V = \min_{v\in\text{supp}(Q_V)} Q(v) \tag{15}$$

$$\mu_{UV} = \min_{(v,u)\in\text{supp}(Q_{UV})} Q(u,v) \tag{16}$$

$$\nu_{V|U} = \max_{(v,u)\in\text{supp}(Q_{UV})} Q(v|u). \tag{17}$$

If $\frac{\nu_{V|U}}{\mu_V} < 1$, we have

$$d_2 \le 2|\mathcal{V}|\cdot|\mathcal{U}|\cdot e^{-2n\epsilon^2\mu_{UV}^2}\cdot\log 2 \tag{18}$$

and $d_2\to 0$ as $n\to\infty$. If $\frac{\nu_{V|U}}{\mu_V}\ge 1$, we have

$$d_2 \le 2|\mathcal{V}|\cdot|\mathcal{U}|\cdot e^{-2n\epsilon^2\mu_{UV}^2}\cdot n\cdot\log\left(\frac{\nu_{V|U}}{\mu_V}+1\right) \tag{19}$$

and $d_2\to 0$ as $n\to\infty$.

Combining the above we have

$$\text{E}[D(P_{V^n}||Q_V^n)] \to 0 \tag{20}$$

if $R > I(V;U)+2\epsilon$ and $n\to\infty$. As usual, (20) means that there must exist a code with $D(P_{V^n}||Q_V^n) < \delta$ for any $\delta > 0$ and sufficiently large n. This proves the coding theorem. The converse follows from [1, Theorem 5.2] by removing the normalization factor $\frac{1}{n}$. ∎

Remark 1: Theorem 1 is proved only for discrete and finite random variables. However, extensions to continuous random variables should be possible.

Remark 2: The cardinality bound on $\text{supp}(Q_U)$. can be derived using techniques from [6, Ch. 15].

Remark 3: If $V = U$, then we have $R > H(V)$.

Theorem 1 is proven using a uniform W which represents strings of uniform bits. If we use a non-uniform W for the coding scheme in Theorem 1, can we still drive the unnormalized informational divergence to zero? We give the answer in the following lemma.

Lemma 1: Let $W = B^{nR}$ be a bit stream with nR bits that are generated i.i.d. with a binary distribution P_X with $P_X(0) = p$, $0 < p \le \frac{1}{2}$. The rate R is achievable if

$$R > \frac{I(V;U)}{H_2(p)} \tag{21}$$

where $H_2(\cdot)$ is the binary entropy function.

Proof: The proof is given in the Appendix. ∎

Remark 4: Lemma 1 states that even if W is not uniformly distributed, the informational divergence can be made small. This is useful because if the distribution of W is not known exactly, then we can choose R large enough to guarantee the desired resolvability result.

IV. Discussion

Hayashi studied the resolvability problem using unnormalized divergence and he derived bounds for nonasymptotic cases [3, Lemma 2]. Theorem 1 can be derived by extending [3, Lemma 2] to asymptotic cases and it seems that such a result was the underlying motivation for [3, Lemma 2]. Unfortunately, Theorem 1 is not stated explicitly in [3] and the ensuing asymptotic analysis was done for *normalized* informational divergence. Hayashi's proofs (he developed two approaches) were based Shannon random coding.

Theorem 1 implies [1, Theorem 6.3] which states that for $R > I(V;U)$ the normalized divergence $\frac{1}{n}D(P_{V^n}||Q_V^n)$ can be made small. Theorem 1 implies [2, Theorem 4] for product distributions through Pinkser's inequality [7, Lemma 11.6.1]

$$D(P_X||Q_X) \ge \frac{1}{2\ln 2}||P_X - Q_X||_{\text{TV}}^2 \tag{22}$$

where

$$||P_X - Q_X||_{\text{TV}} = \sum_x |P(x)-Q(x)|. \tag{23}$$

Moreover, the speed of decay in (13), (18) and (19) is (almost) exponential with n. We can thus make

$$\alpha(n)\cdot\text{E}\left[D(P_{V^n}||Q_V^n)\right] \tag{24}$$

vanishingly small as $n\to\infty$, where $\alpha(n)$ represents a *sub-exponential* function of n that satisfies,

$$\lim_{n\to\infty}\frac{n\cdot\alpha(n)}{e^{\beta n}} = 0 \tag{25}$$

where β is positive and independent of n (see also [3]). For example, we may choose $\alpha(n) = n^m$ for any integer m.

Since all achievability results in [8] are based on [2, Theorem 4], Theorem 1 extends the results in [8] as well. Theorem 1 is further closely related to *strong* secrecy [9] and provides a simple proof that Shannon random coding suffices to drive an *unnormalized* mutual information between messages and eavesdropper observations to zero.

Theorem 1 is valid for approximating product distributions only. However extensions to a broader class of distributions, e.g., *information stable* distributions [2], are clearly possible.

Finally, an example code is as follows (courtesy of F. Kschischang). Consider a channel with input and output alphabet the 2^7 binary 7-tuples. Suppose the channel maps each input uniformly to a 7-tuple that is distance 0 or 1 away, i.e., there are 8 channel transitions for every input and each transition has probability $\frac{1}{8}$. A simple "modulation" code for this channel is the $(7,4)$ Hamming code. The code is perfect and if we choose each codeword with probability $\frac{1}{16}$, then the output V^7 of the channel is uniformly distributed over all 2^7 values. Hence $I(V;U)=4$ bits suffice to "approximate" the product distribution (here there is no approximation).

APPENDIX A
NON-UNIFORM W

Observe that $H(W)=H(B^{nR})=nR\cdot H_2(p)$. Following the same steps as in (10) we have

$$
\begin{aligned}
&\mathrm{E}[D(P_{V^n}||Q_V^n)] = \mathrm{E}\left[\log \frac{\sum_{j=1}^M P(j)Q_{V^n|U^n}(V^n|U^n(j))}{Q_V^n(V^n)}\right] \\
&= \sum_w P(w)\cdot \mathrm{E}\left[\log \frac{\sum_{j=1}^M P(j)Q_{V|U}^n(V^n|U^n(j))}{Q_V^n(V^n)}\middle| W=w\right] \\
&\le \sum_w P(w)\cdot \mathrm{E}\left[\log\left(\frac{P(w)Q_{V|U}^n(V^n|U^n(w))}{Q_V^n(V^n)}+1-P(w)\right)\right] \\
&\le \sum_w P(w)\cdot \mathrm{E}\left[\log\left(\frac{P(w)Q_{V|U}^n(V^n|U^n(w))}{Q_V^n(V^n)}+1\right)\right] \\
&= d_1+d_2+d_3
\end{aligned} \tag{26}
$$

where

$$
\begin{aligned}
d_1 &= \sum_{w\in T_\epsilon^n(P_X^n)} P(w) \sum_{(u^n(w),v^n)\in T_\epsilon^n(Q_{UV}^n)} Q_{UV}^n(u^n(w),v^n) \\
&\quad \left[\log\left(\frac{P(w)Q_{V|U}^n(v^n|u^n(w))}{Q_V^n(v^n)}+1\right)\right] \\
d_2 &= \sum_{w\in T_\epsilon^n(P_X^n)} P(w) \sum_{(u^n(w),v^n)\notin T_\epsilon^n(Q_{UV}^n)} Q_{UV}^n(u^n(w),v^n) \\
&\quad \left[\log\left(\frac{P(w)Q_{V|U}^n(v^n|u^n(w))}{Q_V^n(v^n)}+1\right)\right] \\
d_3 &= \sum_{w\notin T_\epsilon^n(P_X^n)} P(w) \sum_{(u^n(w),v^n)} Q_{UV}^n(u^n(w),v^n) \\
&\quad \left[\log\left(\frac{P(w)Q_{V|U}^n(v^n|u^n(w))}{Q_V^n(v^n)}+1\right)\right].
\end{aligned} \tag{27}
$$

We can bound d_1 to d_3 as follows

$$
\begin{aligned}
d_1 &\le \sum_{w\in T_\epsilon^n(P_X^n)} P(w)\left[\log\left(\frac{2^{n(I(V;U)+2\epsilon)}}{2^{n(R\cdot H_2(p)-\epsilon)}}+1\right)\right] \\
&\le \log\left(2^{-n(R\cdot H_2(p)-I(V;U)-3\epsilon)}+1\right) \\
&\le \log(e)\cdot 2^{-n(R\cdot H_2(p)-I(V;U)-3\epsilon)}
\end{aligned} \tag{28}
$$

which goes to zero if $R>\frac{I(V;U)+3\epsilon}{H_2(p)}$ and $n\to\infty$. We also have

$$
\begin{aligned}
d_2 &\le \sum_{w\in T_\epsilon^n(P_X^n)} P(w) \sum_{(u^n(w),v^n)\notin T_\epsilon^n(Q_{UV}^n)} Q_{UV}^n(u^n(w),v^n) \\
&\quad \left[\log\left(\left(\frac{(1-p)\cdot\nu_{V|U}}{\mu_V}\right)^n+1\right)\right] \\
&\le 2|\mathcal{V}|\cdot|\mathcal{U}|\cdot e^{-2n\epsilon^2\mu_{UV}^2}\log\left(\left(\frac{\nu_{V|U}}{\mu_V}\right)^n+1\right)
\end{aligned} \tag{29}
$$

which goes to zero as $n\to\infty$ (see (14)). We further have

$$
\begin{aligned}
d_3 &\le \sum_{w\notin T_\epsilon^n(P_X^n)} P(w) \sum_{(u^n(w),v^n)\in T_\epsilon^n(Q_{UV}^n)} Q_{UV}^n(u^n(w),v^n) \\
&\quad \left[\log\left(\left(\frac{(1-p)\cdot\nu_{V|U}}{\mu_V}\right)^n+1\right)\right] \\
&\le \sum_{w\notin T_\epsilon^n(P_X^n)} P(w)\left[\log\left(\left(\frac{(1-p)\cdot\nu_{V|U}}{\mu_V}\right)^n+1\right)\right] \\
&\le 4\cdot e^{-2n\epsilon^2p^2}\log\left(\left(\frac{\nu_{V|U}}{\mu_V}\right)^n+1\right)
\end{aligned} \tag{30}
$$

which goes to zero as $n\to\infty$ (see (18) and (19)).

Combining the above for non-uniform W we have

$$
\mathrm{E}[D(P_{V^n}||Q_V^n)] \to 0 \tag{31}
$$

if $R>\frac{I(V;U)+3\epsilon}{H_2(p)}$ and $n\to\infty$.

ACKNOWLEDGMENT

J. Hou and G. Kramer were supported by an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research. G. Kramer was also supported by NSF Grant CCF-09-05235. The authors thank G. Böcherer, F. Kschischang and M. Bloch for useful remarks.

REFERENCES

[1] A. Wyner. The common information of two dependent random variables. *IEEE Trans. Inf. Theory*, 21(2):163–179, March 1975.

[2] T.S. Han and S. Verdú. Approximation theory of output statistics. *IEEE Trans. Inf. Theory*, 39(3):752–772, May 1993.

[3] M. Hayashi. General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel. *IEEE Trans. Inf. Theory*, 52(4):1562–1575, April 2006.

[4] J.L. Massey. *Applied Digital Information Theory*. ETH Zurich, Zurich, Switzerland, 1980-1998.

[5] A. Orlitsky and J.R. Roche. Coding for computing. *IEEE Trans. Inf. Theory*, 47(3):903–917, March 2001.

[6] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic, New York, 1981.

[7] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, NY, USA, 2006.

[8] P.W. Cuff, H.H. Permuter, and T.M. Cover. Coordination capacity. *IEEE Trans. Inf. Theory*, 56(9):4181–4206, Sept. 2010.

[9] M. Bloch and J. Barros. *Physical Layer Security From Information Theory to Security Engineering*. Cambridge University Press, 2011.